

# THE INFLUENCE OF GRAPHICAL AND SYMBOLIC LANGUAGE MANIPULATIONS ON RESPONSES TO SELF-ADMINISTERED QUESTIONS

LEAH MELANI CHRISTIAN  
DON A. DILLMAN

**Abstract** This article reports results from 14 experimental comparisons designed to test 7 hypotheses about the effects of two types of nonverbal languages (symbols and graphics) on responses to self-administered questionnaires. The experiments were included in a survey of 1,042 university students. Significant differences were observed for most comparisons, providing support for all seven hypotheses. These results support the view that respondents' answers to questions in self-administered surveys are influenced by more than words. Thus, the visual presentation of questions must be taken into consideration when designing such surveys and, especially, when comparing results across surveys in which the visual presentation of questions is varied.

It has been recognized for many years that answers to self-administered questionnaires are influenced by the ways in which questions and answers are displayed on questionnaire pages (e.g., Rothwell, 1985; Smith 1993, 1995; Wright and Barnard 1975, 1978). However, our scientific understanding of the nature of these effects is not well developed. Although it has been argued on theoretical grounds that visual layout and design make a difference in how people answer questionnaires (Jenkins and Dillman 1997; Sless 1994), more experimental evidence is needed to understand how changing

LEAH MELANI CHRISTIAN is a graduate research assistant in the Social and Economic Sciences Research Center (SESRC) and the Department of Sociology at Washington State University (WSU), and DON A. DILLMAN is the Thomas S. Foley Distinguished Professor of Government and Public Policy and Regents' Professor in the WSU Departments of Sociology and Community and Rural Sociology, and Deputy Director of the SESRC, Pullman, Washington 99164-4014. The authors wish to acknowledge the financial support provided for this research by the WSU Agricultural Research Center under Western Region Project W-183, the SESRC, the National Science Foundation Division of Science Resource Statistics, the U.S. Department of Agriculture National Agricultural Statistics Service, and the Gallup Organization. Appreciation is also expressed to Thom Allen, who served as study director for collection of the data analyzed here. Address correspondence to Don A. Dillman; e-mail: dillman@wsu.edu.

the visual presentation of individual survey questions influences people's answers.

In contrast to interviews, which rely mostly on verbal language for presenting questions to respondents, questions in self-administered questionnaires are also presented by means of several nonverbal languages. These nonverbal languages include symbolic language (the use of symbols with shared cultural meaning) and graphical language (the use of multiple design features such as font size, brightness, location, and spacing) that may convey certain meanings, apart from those conveyed solely by words. Our purpose in this article is to report results focusing on these nonverbal languages, which have received little attention. The study involved six independent manipulations and one combined manipulation of graphical and symbolic languages to determine whether answers to survey questions were affected by these manipulations in a self-administered, paper questionnaire. All experiments were included in a survey of university students designed to obtain an evaluation of their student experience at Washington State University.

## Theoretical Background

Respondents act as cooperative communicators, and they will endeavor to make sense of survey questions by drawing on all information given by the researcher (Schwarz 1996, p. 41). "In a research situation, the researcher's contributions include apparently formal features of the questionnaire, such as the numeric values of rating scales or the scales' graphical layout" (Schwarz, Grayson, and Knäuper 1998, p. 182). Especially when respondents are unsure about what is being asked or how to report their answer, they will draw information from the context of the interview conversation or, in self-administered questionnaires, from the context of the question. Schwarz (1996, p. 48) says, "[T]o disambiguate its meaning respondents turn to the context of the question, much as they would in daily life," and this context includes the formal features or nonverbal languages of the questionnaire. Thus, while words make up an important source of question-meaning for respondents, so too do formal features or nonverbal languages used in self-administered questionnaires.

### VERBAL LANGUAGE

Self-administered questionnaires consist of information presented in four distinct languages: verbal, numerical, graphical, and symbolic. These languages can independently and jointly influence respondent behavior. Verbal language is used in all modes of data collection and as such is very important in survey design. A great amount of research in survey methodology has focused on the importance of carefully choosing words to convey the

researcher's meaning (e.g., Payne 1951; Schuman and Presser 1981; Sudman and Bradburn 1974). In self-administered questionnaires, the additional non-verbal languages (numerical, graphical, and symbolic) may affect whether questions are read, the order in which they are read, and the meaning conveyed to respondents.

#### NUMERICAL LANGUAGE

Research has shown that numerical language independently influences respondent behavior (Schwarz et al. 1985; Schwarz et al. 1991). Schwarz et al. (1985) showed that respondents gain information about the researcher's expectations from the response alternatives in surveys—using the verbal and numeric labels of ordinal scales as frames of reference. Schwarz et al. (1991) manipulated the numeric values attached to the endpoints of opinion scales and found that this independent numerical language change affected respondents' answers. Thus, respondents may use numerical language as a source of question-meaning above and beyond the verbal labels of the scale.

#### GRAPHICAL LANGUAGE

Graphical language includes features such as size, brightness and color, shape, location, and spatial arrangement of words, numbers, and symbols. Graphical language acts like "paralanguage," optional voice effects that accompany the sounds of an utterance and may convey meaning, used in aural communication. Thus, graphical language includes features of the questionnaire that accompany words, numbers, and symbols and convey meaning. The verbal, numerical, and symbolic languages are transmitted through the visual channel via graphic paralanguage (Redline and Dillman 2002, p. 181). For example, when a respondent reads a question, the graphical language can vary the stimulus seen by the respondent by changing the size of words, the color of symbols, or the graphical layout of a scale. These changes, in turn, can influence respondents' answers.

Smith (1993, 1995) reviewed tests of three independent, accidental graphical manipulations in which the location of answer spaces, the size of answer spaces in an open-ended question, and the size of boxes used to display a socioeconomic scale were all varied. In the first manipulation, a font error misaligned the response categories in the second part of a three-item question. In this version, the "yes" box was placed in the "no" column, and the "no" box was further out to the right. Analyzing the response distributions, it is apparent that some respondents confused the "yes" and "no" boxes because of the error in the location of the response boxes.

In the other two accidental manipulations of graphical features, Smith (1993) observed that when respondents were given a larger answer space for

an open-ended question, they wrote longer answers. He also observed that changing the layout of the scale from a stack of ten equal-sized vertical boxes reflecting a ladder shape to a stack of ten different-sized vertical boxes reflecting a pyramid shape altered the response distribution because respondents appeared to gain information about how the researcher thought socioeconomic status was distributed from the graphical layout of the scale. Schwarz, Grayson, and Knäuper (1998) extended Smith's research on the socioeconomic status question by including an item about academic performance on a survey given to a university population. Some respondents were given the ten vertically stacked boxes, some were given the pyramid version, and some were given a new test where the boxes replicated an onion shape—the middle boxes were larger, and the top and bottom boxes were smaller. Respondents rated their academic performance less favorably on the pyramid version (more respondents chose the wider boxes).

Schwarz (1996) discusses unpublished research by Schwarz and Hippler, who in 1992 conducted a study manipulating one aspect of graphical language by grouping two questions (about general satisfaction and marital satisfaction) together in one box, and they found that respondents appeared more likely to see the two questions as related. However, when the questions were separated, respondents appeared to treat them as separate questions; the correlation between responses to the two questions was reduced when they were graphically separated.

#### SYMBOLIC LANGUAGE

Symbolic language uses signs that have cultural meaning to convey information to respondents. For example, an arrow is used to tell a person to go in the direction the arrow is pointing. In comparison to the research examined above, there is relatively little systematic research on symbolic language effects. Some experimental evidence shows that combined symbolic, graphical, and verbal language changes reduced branching errors significantly in self-administered questionnaires. The newly designed versions simultaneously manipulated the font size of the branching instruction, use of directional arrows, graphical spacing, and additional verbal language to help respondents both prevent and detect their errors. In a student classroom experiment in which 1,266 students completed one of three versions of a questionnaire, commission errors (not skipping ahead when directed to do so) were reduced from 20 percent to between 7 and 9 percent for the newly designed instructions (Redline and Dillman 2002). In a follow-up test embedded in the 2000 decennial census of the United States, commission errors were reduced from 21 percent to 13–15 percent for two similar methods (Redline et al. 2003). A shortcoming of both of these experiments is that the independent effects of the verbal, symbolic, and graphical language changes were impossible to disentangle. Some experimental manipulations reported here focus exclusively on potential symbolic design effects.

### Procedure

Fourteen experiments were embedded on pages 2, 3, and 4 of a four-page questionnaire developed for assessing the student experience at Washington State University and conducted from March to April 2001. It was printed in a two-column format on 8½-by-11-inch pages, with a colored background used to contrast with the white answer spaces provided for both open- and closed-ended questions.

Four versions of the questionnaire were mailed to equal subsamples (450) of a random sample of 1,800 undergraduate students living in the Pullman, Washington area (students on other campuses or enrolled in the distance degree program were excluded). The experimental questions reported here were the same in two of the four versions of the questionnaires.<sup>1</sup> A \$2 incentive was enclosed with the first mailing, and a follow-up postcard and one replacement questionnaire were mailed. Response rates were calculated using Response Rate 1 (RR1—according to American Association for Public Opinion Research [AAPOR] definitions, the number of completed questionnaires divided by the total number of questionnaires mailed) and were virtually the same for both versions, 57.7 percent (519 of 900) and 58.1 percent (523 of 900), obtaining a total response rate of 57.9 percent (1,042 of the 1,800 questionnaires mailed).

The inclusion of so many experiments in one questionnaire raises issues of whether some of the experiments may have affected results for others. That possibility cannot be ruled out. For the most part, however, the alternative treatments tested here are not ones that seem likely to influence respondents in such a way that answers to subsequent questions would somehow be affected. It is quite possible that percentage distributions to individual questions were influenced through order effects, but we were unable to find evidence of such effects.

Statistical tests to evaluate the hypotheses include chi-square tests for differences and *t*-tests for mean differences, where appropriate. The tests conducted vary depending upon the questions and therefore will be reported for each hypothesis.

### Graphical Language Experiments

#### PLACEMENT OF INSTRUCTION

Four independent graphical language manipulations were designed to determine the effects on response behavior. In the first manipulation, graphical

1. Four versions were needed for three other experiments involving the construction of pages 1 and 3 and have been reported elsewhere (Sawyer and Dillman 2002). All of the experiments reported here were the same on Versions A and D and Versions B and C.

location was varied by placing a special instruction before and after the response categories (figure 1). The location before the response categories places the instruction in the navigational path, so that respondents can process the instruction after reading the query but before a response is required. It has been argued that an important goal of questionnaire construction is to use graphical language in a way that makes the elements (number, query, instructions, and answer choices) of a question appear as a distinct group, and

### Graphical Language Changes

#### *Location of special instruction before vs. after answer categories.*

- |  |  |
|--|--|
| <p>A. Have one-on-one meetings with professors contributed significantly to your WSU education?</p> <p>If you haven't had many one-on-one meetings, just skip to Question 9.</p> <p><input type="checkbox"/> Yes<br/><input type="checkbox"/> No</p> | <p>B. Have one-on-one meetings with professors contributed significantly to your WSU education?</p> <p><input type="checkbox"/> Yes<br/><input type="checkbox"/> No</p> <p>If you haven't had many one-on-one meetings, just skip to Question 9.</p> |
|--|--|

#### *Linear vs. nonlinear scale layout.*

- |   |  |
|---|--|
| <p>A. How satisfied are you with the variety of classes offered for students at WSU?</p> <p><input type="checkbox"/> Very Satisfied    <input type="checkbox"/> Somewhat Dissatisfied<br/><input type="checkbox"/> Somewhat Satisfied    <input type="checkbox"/> Very Dissatisfied</p> | <p>B. How satisfied are you with the variety of courses offered at WSU?</p> <p><input type="checkbox"/> Very Satisfied<br/><input type="checkbox"/> Somewhat Satisfied<br/><input type="checkbox"/> Somewhat Dissatisfied<br/><input type="checkbox"/> Very Dissatisfied</p>       |
| <p>A. Overall, how would you rate the quality of education that you are getting at WSU?</p> <p><input type="checkbox"/> Excellent    <input type="checkbox"/> Good    <input type="checkbox"/> Poor<br/><input type="checkbox"/> Very good    <input type="checkbox"/> Fair</p>         | <p>B. Overall, how would you rate the quality of education that you are getting at WSU?</p> <p><input type="checkbox"/> Excellent<br/><input type="checkbox"/> Very good<br/><input type="checkbox"/> Good<br/><input type="checkbox"/> Fair<br/><input type="checkbox"/> Poor</p> |

#### *Equal vs. unequal spacing between response boxes.*

- |  |  |
|--|--|
| <p>A. What percentage of WSU students do you think use the Student Recreation Center?<sup>1</sup></p> <p><input type="checkbox"/> Less than one-fourth of WSU students<br/><input type="checkbox"/> About half of WSU students<br/><input type="checkbox"/> About three-fourths of WSU students<br/><input type="checkbox"/> More than three-fourths of WSU students<br/><input type="checkbox"/> No opinion</p> | <p>B. What percentage of WSU students do you think use the Student Recreation Center?<sup>1</sup></p> <p><input type="checkbox"/> Less than one-fourth of WSU students<br/><input type="checkbox"/> About half of WSU students<br/><input type="checkbox"/> About three-fourths of WSU students<br/><input type="checkbox"/> More than three-fourths of WSU students<br/><input type="checkbox"/> No opinion</p> |
| <p>A. Thinking about your life after completing your education, which one of these do you consider most important?</p> <p><input type="checkbox"/> To have a life partner with whom you have a satisfying relationship<br/><input type="checkbox"/> Enjoy your work<br/><input type="checkbox"/> Earn a high income<br/><input type="checkbox"/> Raise a family</p>  | <p>B. Thinking about your life after completing your education, which one of these do you consider most important?</p> <p><input type="checkbox"/> To have a life partner, with whom you have a satisfying relationship<br/><input type="checkbox"/> Enjoy your work<br/><input type="checkbox"/> Earn a high income<br/><input type="checkbox"/> Raise a family</p>   |

*Use of larger answer spaces on open-ended answers.*

A. Why did you choose to attend Washington State University?

B. Why did you choose to attend Washington State University?

A. In your own words how would you describe your advisor(s)?

B. In your own words how would you describe your advisor(s)?

A. What additional recreational activities, if any, would you like to see offered at the Student Recreation Center?

B. What additional recreational activities, if any, would you like to see offered at the Student Recreation Center?

<sup>1</sup>The query was displayed on one continuous line.

**Figure 1.** Question formats used to test hypotheses of graphical language manipulations.

instructions should be provided exactly where they are needed by the respondents (Dillman 2000, pp. 96–99).

In this experiment a yes/no question was developed that was expected not to apply to a number of respondents.<sup>2</sup> In one version, the special instruction to skip to the next question when applicable was placed ahead of the response categories, and in the other version it was placed immediately after the categories. It was expected that more people would skip the question when the instructions were located before the response options because they were more likely to see and thus read the instructions before answering. Thus, it is hypothesized that special instructions are more likely to be followed correctly

2. A filter question was not used because this experiment was specifically testing the location of special instructions within or outside of the navigational path.

if they are located in the navigational path just before, rather than just after, the place of their intended use.

#### LINEAR VERSUS NONLINEAR LAYOUT

A second graphical manipulation compared two linear ordinal scales (where all scale points were listed vertically) to a nonlinear layout (where responses were triple and double banked). Using an ordinal scale, respondents must decide where to place themselves on an implied continuum. Graphically speaking, a linear layout of ordinal scales would seem to facilitate the process of identifying where a respondent best fits on the continuum. When the responses are viewed in a nonlinear layout, the graphical language conveying the 4 or 5-point scale is interrupted. Nonlinear layouts such as double- and triple-banking responses are commonly used to save space, but the potential implications for responses are usually not considered. Two experiments were conducted, linear layout versus triple banking of an ordinal scale and linear layout versus double banking of an ordinal scale (figure 1). It is hypothesized that the percentage of respondents choosing categories from the top line in the nonlinear format will be significantly greater than the percentage selecting from the lower lines, and in particular the percentage of respondent choosing the category just to the right of the first category will increase in the nonlinear layout because some respondents may read the top line only, ignoring the bottom line.

#### CATEGORY SPACING

Graphical location and spatial arrangement can also be manipulated by increasing the space between response categories. No rules appear to exist for how far apart response categories should be placed from one another. Intuitively, it would seem that response categories should be placed equal distances from one another. However, the ramifications of unequal distances between categories are still unclear. As mentioned earlier, Schwarz (1996) noted that respondents use response categories to gain information about how to report their answers. In one experiment, we examine spacing of categories in a nominal scale where the stem of the question reads, "Thinking about your life after completing your education, which one of these do you consider most important?" (figure 1). Because the stem directs the respondent to the answer categories to determine what is important after graduating, respondents are expected to learn about the specific aim of the question from the response categories themselves. In the second experiment, we examine an ordinal scale where respondents simply place themselves in a category, but they know from the stem that the question is asking for a percentage. In both of these manipulations, we are testing whether unequal spacing will increase the visual prominence of some categories and result in more respondents choosing them.



Our hypothesis is that the spacing manipulation will lead respondents to select categories that are set off from others more frequently, and that this tendency will be especially evident in the case of the question with nominal categories. In a previous branching instruction experiment (Dillman and Carley-Baxter 1999) we found that 3 out of 52 questions produced significantly different responses among treatment groups. Two of the three questions (one using a nominal scale and the other an ordinal scale) produced significant results because of the greater selection of one response option when it was graphically separated from the other options. In the versions where these responses were chosen more often, answer boxes were placed to the right of response options. The possibility that the relationship between the answer box and the category label influenced responses in some unknown way suggested the need for testing the spacing of the response options independently of the placement of the answer boxes.

#### SIZE OF ANSWER SPACES

The final examination of graphical effects involved three experiments testing how increasing the size of open-ended answer spaces would affect respondent behavior. The size of the open-ended answer space was doubled on one version of each of three questions (figure 1). In self-administered questionnaires where the query and response spaces are the only information given to the respondents and further probing by interviewers cannot be done, it has been shown that respondents typically provide shorter, less complete answers to open-ended questions than when surveyed in interviews (Dillman 2000). Based upon work by Stember (1956), Smith (1993) has argued that allowing more space for recording open-ended answers in interview-administered surveys produces longer recorded responses that may be closer to actual verbatim. It is hypothesized that providing a larger answer space will produce longer answers with more words. This experiment extends the previous research by Smith (1993) to determine whether longer answers mean respondents substantively report more themes or whether respondents use more words to provide more detail in their responses.

#### FINDINGS

*Placement of Instruction.* This graphical location change dramatically affected whether respondents used the special instructions. When the special instruction "If you haven't had many one-on-one meetings, just skip to Question 9" was placed after the response options, 54.9 percent of the respondents said "yes," 40.3 percent said "no," and 4.8 percent provided no answer (table 1). However, when these instructions were placed before the responses, 54.7 percent said "yes," 19.1 percent said "no," and 26.2 percent provided no

**Table 1.** Independent Graphical Language Manipulations: Percentage of Respondents Choosing Each Response When Instruction to Skip Question Was Located Before versus After the Response Options

Responses	Instructions before (N=519)	Instructions after (N=523)
Yes	54.7	54.9
No	19.1	40.3
Missing	26.2	4.8
Total	100	100

NOTE.—Answered versus Missing:  $\chi^2 = 106.67$ ,  $p = .000$ ; Yes versus No versus Missing:  $\chi^2 = 116.99$ ,  $p = .000$ .

answer ( $\chi^2 = 116.99$ ,  $p = .000$ ; “yes” versus “no” versus missing, by version). Thus, when special instructions were graphically placed where respondents were more likely to see them, prior to answering the question, the location influenced their decision to answer or not answer the question. Not only did placing the instructions after the response choices produce different results on this question, but it also introduced confusion, as some respondents appear to have used the instructions for the following question as well. Item non-response to the following question increased from 2.6 percent to 11.0 percent when the instructions were placed after the responses.

*Linear versus Nonlinear Layout.* Changing the scale layout from a linear to nonlinear layout affected respondent behavior in one of the two experiments. Respondents were significantly more likely ( $\chi^2 = 6.66$ ,  $p = .010$ ) to select responses from the top line in the nonlinear version (see table 2). Specifically, 40.4 percent of respondents chose “good” and 42.4 percent chose “very good” on the nonlinear version. In the linear format, 31.3 percent of respondents chose “good” and 48.8 percent chose “very good.” This difference by version between “good” and “very good” is significant ( $\chi^2 = 8.2$ ,  $p = .004$ ). More respondents chose “good” and fewer chose “very good” when the scale was triple banked, suggesting perhaps that some respondents focused only on the top line.

A slight trend in the same direction exists for the second question, which used double banking, but the chi-square ( $\chi^2 = .16$ ,  $p = .685$ ; top versus bottom line) is not significant (nor are the individual chi-squares for differences between individual answer categories). Here, the two horizontal categories “very satisfied” and “somewhat dissatisfied” do not display the sense of a complete scale, as might be inferred from the three horizontal categories—“excellent,” “good,” “poor”—on the question for which differences are statistically significant. Thus, differences based on content (verbal language) and the number of response options between the two tests of this hypothesis could have produced the difference in results.

**Table 2.** Independent Graphical Language Manipulations: Percentage of Respondents Choosing Each Response for Linear versus Nonlinear Graphical Layout

Responses	Question 1		Question 2	
	Linear (N=518)	Nonlinear (N=517)	Responses	Linear (N=519) Nonlinear (N=518)
(1) Excellent	11.4	11.0	(1) Very satisfied	42.0 38.8
(2) Very good	48.8	42.4	(2) Somewhat satisfied	49.5 51.2
(3) Good	31.3	40.4	(3) Somewhat dissatisfied	6.9 8.9
(4) Fair	7.0	5.4	(4) Very dissatisfied	1.5 1.2
(5) Poor	1.5	.8		
Total	100	100		100 100
Mean	2.4	2.4		1.7 1.7
Difference of Means	$t = .8, p = .206$			$t = 1.1, p = .146$
Overall Chi-square	$\chi^2 = 10.8, p = .029$			$\chi^2 = 2.3, p = .509$

NOTE.—Question 1: Good versus Very good:  $\chi^2 = 8.2, p = .004$ ; Top line (1, 3, 5) versus Bottom line (2, 4):  $\chi^2 = 6.7, p = .010$ . Question 2: Top line (1, 3) versus Bottom line (2, 4):  $\chi^2 = .16, p = .685$ . No significant differences for individual responses.

*Category Spacing.* The effect of equal versus unequal spacing between response categories was significant ( $\chi^2=4.8$ ,  $p=.028$ ; unequally spaced response versus all others) for the nominal scale question but not the ordinal scale question ( $\chi^2=1.4$ ,  $p=.844$ ; overall chi-square) (table 3). The content of the response options in the nominal scale question were more likely to give information about the aim of the question. The response option, "To have a life partner with whom you have a satisfying relationship," was selected significantly more times on the unequally spaced version (37.6 percent versus 30.8 percent). However, in the ordinal scale question, the respondent knows what the researcher is looking for from the question stem. We suspect this question was less susceptible to graphical spacing effects because less meaning is gained from the response categories than in the nominal scale question.

*Size of Answer Spaces.* Varying the amount of answer space on open-ended questions influenced both the number of words and the number of themes provided by respondents (table 4). The number of words was hand-counted for each open-ended response, and themes were counted as the number of topics that were mentioned. The coding of themes was completed by one researcher, and 10 percent were verified by another researcher with 90 percent agreement (two researchers coded 10 percent of the themes independently, and 90 percent of the time they coded the same number of themes). For all three questions, the larger space produced longer answers with a significantly greater number of words: 13.3 versus 9.7 ( $t=6.5$ ,  $p=.000$ ), 12.9 versus 6.6 ( $t=12.9$ ,  $p=.000$ ), and 12.0 versus 10.2 ( $t=1.8$ ,  $p=.039$ ). In two of the three questions, the larger answer space significantly produced a greater number of themes or topics mentioned in the answer: 2.0 versus 1.8 ( $t=2.7$ ,  $p=.003$ ) and 2.1 versus 1.7 ( $t=8.0$ ,  $p=.000$ ). (The third question was 1.5 versus 1.4 with  $t=.7$ ,  $p=.195$ .) These results support earlier findings that larger answer spaces for open-ended questions produce longer answers. We also find in this case that the longer answers generally contain more topics.

In summary, the experiments examining graphical features of self-administered questionnaires found that placement of instructions, layout and spacing of categories, and size of answer spaces influenced response behavior. We turn now to experiments on symbol usage and a joint graphical/symbol manipulation.

## Symbolic Language Experiments

### LINES IN ANSWER SPACES

The first symbolic language manipulation tested the effects of adding lines in open-ended answer spaces. We expected that respondents would write on the lines, and that the lines would further indicate how the respondents should

**Table 3.** Independent Graphical Language Manipulations: Percentage of Respondents Choosing Each Response with Equal versus Unequal Spacing between Response Categories

Responses	Nominal Scale		Ordinal Scale	
	Equal (N=474)	Unequal (N=471)	Equal (N=510)	Unequal (N=514)
To have a life partner with whom you have a satisfying relationship <sup>a</sup>	30.8	37.6	17.3	19.1
Enjoy your work	43.9	42.9	47.5	46.5
Earn a high income	10.6	7.9	24.7	22.6
Raise a family	14.8	11.7	4.3	4.7
			6.3	7.2
Total	100	100	100	100
Overall Chi-square	$\chi^2 = 6.8, p = .079$		$\chi^2 = 1.4, p = .844$	

NOTE.—Nominal Scale, "To have a life partner" versus all others:  $\chi^2 = 4.8, p = .028$ . Ordinal Scale, means: 2.3, 2.3; difference of means:  $t = .07, p = .472$ . No significant differences for response option 1 or no opinion when run separately or together.

<sup>a</sup> These answer boxes were separated by one or two lines from the other responses.

**Table 4.** Independent Graphical Language Manipulations: Mean Number (#) of Words and Themes Reported by Respondents in Larger versus Smaller Answer Spaces

Question	Mean #	Larger space	Smaller space	Difference of Means
Why attend WSU <sup>a</sup>	Words	13.3	9.7	$t=6.5$ $p=.000$
	Themes	2.0	1.8	$t=2.7$ $p=.003$
Description of advisor <sup>b</sup>	Words	12.9	6.6	$t=12.9$ $p=.000$
	Themes	2.1	1.7	$t=8.0$ $p=.000$
Add'l recreational activities <sup>c</sup>	Words	12.0	10.2	$t=1.8$ $p=.039$
	Themes	1.5	1.4	$t=0.7$ $p=.195$

<sup>a</sup> N=500 for Larger space, 509 for Smaller space.

<sup>b</sup> N=484 for Larger space, 494 for Smaller space.

<sup>c</sup> N=284 for Larger space, 280 for Smaller space.

report their answers (just as the size of the answer boxes also conveyed the researcher's expectations for more detailed answers). The first question asked what additional stores or businesses the respondent would like to see in the area (see figure 2). Here, the answer space is small, and adding lines only gives the respondent two lines on which to write, suggesting that only a few words should be used to respond. On this question, we expect that when given only two lines, respondents will report fewer stores or businesses.

The second question asked the respondent to report one or two changes that would improve the educational experience at the university. The answer space provided was quite large (figure 2); thus it would seem that respondents should provide more detail about the one or two changes that they recommend (indicating there should be no significant difference in the number of themes) and suggesting that respondents should answer using prose or paragraph format. It is hypothesized that especially on the lined version, where twelve lines are provided, respondents will follow the researcher's expectations of using prose and providing more detailed, longer responses.

#### DIRECTIONAL ARROW

The second symbolic manipulation was the addition of an arrow to direct respondents toward answering a subordinate question. Sometimes questionnaire designers wish to direct respondents who choose a particular answer to a subordinate question for which another answer is desired. An example, and the one tested here, was embedded in a question that asked where the respondent wanted to live after completing college, "eastern Washington" or "somewhere else" (as shown in figure 2). After the answer choice "somewhere else," the word "where" was listed on the same line as the answer choice, approximately 10 spaces beyond the end of the category description and 26 spaces beyond the answer box. In one version, an arrow was placed between the answer "somewhere else" and the subordinate question, "where."

It has been shown that when respondents read text they focus on a space of about 2 degrees, or 8–10 characters in width (Kahneman 1973). This distance is known as the "foveal view." The arrow was placed so that the subordinate question would be visible in the foveal view. An arrow is a symbol that is culturally defined to focus one's attention in the direction the arrow is pointing. We reasoned that respondents were more likely to see and, as a result, respond to the subordinate question when an arrow was placed between the category description and the subordinate question. The arrow was chosen because it had been employed in a combined manipulation of multiple languages that reduced branching errors in a study by Redline et al. (2003). It is hypothesized that the addition of an arrow between a response category and a subordinate question will increase the likelihood that respondents will answer the subordinate question.

## Combined Graphical and Symbolic Language Experiment

### POLAR POINT SCALE VERSUS ANSWER BOX

The final manipulation was a combined graphical and symbolic manipulation in which a polar point scale (labeled endpoints only) was compared to an answer box (where respondents write in the number corresponding to their answer). This experiment tests the effects of eliminating graphical (linear layout of choices) and symbolic language (the check boxes associated with each category) in one version of the question (see figure 2). It has been argued that the use of an answer box might make it possible to provide equivalent stimuli across survey modes because the additional graphical and

#### Symbolic Language Changes

*Addition of lines in open-ended questions.*

- A. What store(s) or business(es) would you most like to see in the Pullman and Moscow area that are currently not available?

- B. What store(s) or business(es) would you most like to see in the Pullman and Moscow area that are currently not available?

- A. If there were one or two changes you could make at this university to improve your educational experience what change would you like to see?

- B. If there were one or two changes you could make at this university to improve your educational experience what change would you like to see?

*Addition of arrow to identify subordinate question.*

- A. After finishing school at Washington State University, where do you hope to live?

☐ Eastern Washington  
☐ Somewhere else → Where? \_\_\_\_\_

- B. After finishing school at Washington State University, where do you hope to live?

☐ Eastern Washington  
☐ Somewhere else → Where? \_\_\_\_\_



## Graphical and Symbolic Language Changes

Polar point scale vs. answer box.

A. On a scale of 1 to 5, with one being very satisfied and 5 being very dissatisfied, how satisfied are you with the classes you are taking this semester?	B. On a scale of 1 to 5, with one being very satisfied and 5 being very dissatisfied, how satisfied are you with the classes you are taking this semester?
<input type="checkbox"/> 1 Very Satisfied <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 Very Dissatisfied	<input type="text"/> Number of your rating
A. On a scale of 1 to 5, where 1 means very satisfied and 5 means not at all satisfied, how do you feel about the quality of <u>instruction</u> in the classes you have taken at WSU?	B. On a scale of 1 to 5, where 1 means very satisfied and 5 means not at all satisfied, how do you feel about the quality of <u>instruction</u> in the classes you have taken at WSU?
<input type="checkbox"/> 1 Very Satisfied <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 Not at all Satisfied	<input type="text"/> Number of your rating
A. On a scale of 1 to 5, where 1 means outstanding and 5 means terrible, how would you rate the quality of advising you have received as a WSU student?	B. On a scale of 1 to 5, where 1 means outstanding and 5 means terrible, how would you rate the quality of advising you have received as a WSU student?
<input type="checkbox"/> 1 Outstanding <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 Terrible	<input type="text"/> Number of your rating

**Figure 2.** Question formats used to test hypotheses of symbolic and combined language manipulations.

symbolic language cannot be provided in aural modes (Dillman 2000, pp. 235–36). Determining the effects of removing these languages within self-administered questionnaires, and thus relying solely on words, is a related concern that needs to be tested before further cross-mode studies are conducted. Three items tested the effect of answer boxes versus linear polar point scales for responding to ordinal scales.

One of the potential difficulties of the number box format is that it requires respondents to remember the specifics of the scale when providing their answer, whereas the inclusion of graphical and numerical information in the polar point scales provides a reminder of how the scale is constructed and is located within the foveal view (8–10 characters) of the response options. As discussed earlier, experiments by Smith (1993, 1995) and Schwarz and Hippler (Schwarz 1996) show that changing the graphical layout of a scale can influence respondent behavior. Since the graphical layout of the scale provides additional support to the verbal and numeric languages in the polar point version, it is hypothesized that the answer box will produce different responses, as the additional graphical and symbolic language support is removed.

## FINDINGS

*Lines in Answer Spaces.* Virtually all respondents understood that they were supposed to use the lines when provided (only 1.2 percent of respondents did not use the lines on the first question and .8 percent on the second question). On the first question, a significantly greater number of words (6.7 versus 5.3;  $t=-4.91$ ,  $p=.000$ ) were used on the unlined version, suggesting that the addition of the lines shortened the appearance of the answer space. However, the difference in means for the themes (2.3 versus 2.4) was not significant ( $t=-.62$ ,  $p=.269$ ), suggesting that instead of reducing the number of stores listed, respondents abbreviated or shortened the name of the stores or businesses, resulting in fewer words used (table 5).

In the case of the second question, neither the number of words nor the number of themes was significantly different across the two versions. More respondents used prose or paragraph format to report their answers on the lined version (48.0 percent) than on the version without lines (42.4 percent); this difference was in the expected direction but not significant ( $\chi^2=2.4$ ,  $p=.120$ ). Further, the addition of lines increased the number of respondents who left the space blank on this question, from 19.4 percent to 26.6 percent ( $\chi^2=7.8$ ,  $p=.005$ ). On the other hand, further analysis indicated that some respondents did not provide an answer to the question and instead wrote responses like "None" or "I have had a wonderful experience at WSU," suggesting that they felt a need to fill the space even when not answering the specific question. This was significantly more likely to occur on the version without lines (7.1 versus 3.1;  $t=7.06$ ,  $p=.008$ ). Thus, the version of the question without lines did not necessarily provide more substantive answers to the question, even though more respondents wrote something in the answer space.

*Directional Arrow.* The addition of the arrow to direct respondents toward a subordinate question significantly increased the percentage of eligible respondents answering the subordinate question (93.9 percent versus 90.8 percent) (table 6). However, one concern about the use of the arrow is that it also increased the ineligible mentions from .5 percent to 2.1 percent. In addition, the arrow decreased the number of missing or blank responses (4.2 percent versus 8.8 percent). The overall chi-square is significant ( $\chi^2=11.65$ ,  $p=.003$ ). Thus, this symbolic language manipulation significantly influenced respondent behavior. The results of this experiment indicate that the use of the arrow did help direct respondents toward a subordinate question and increased the item response to that question; however, researchers must weigh the potential adverse effects of adding an arrow, as it may direct ineligible respondents to the subordinate question as well.

*Polar Point Scale Versus Answer Box.* This final experiment tested the combined manipulation of graphical and symbolic language by replacing the graphical layout of a polar point scale with an answer box in which the respondent is supposed to write in the number corresponding to their answer.

**Table 5.** Independent Symbolic Language Manipulations: Mean Number (#) of Words and Themes Reported by Respondents in Lined versus Unlined Answer Spaces

Question	Mean #	Lines	No Lines	Difference of Means
Add'l stores in Pullman <sup>a</sup>	Words	5.3	6.7	$t = -4.9$ $p = .000$
	Themes	2.3	2.4	$t = -.6$ $p = .269$
Changes for WSU <sup>b</sup>	Words	27.9	28.8	$t = -.5$ $p = .296$
	Themes	1.7	1.7	$t = -.3$ $p = .391$

NOTE.—Only 5 respondents (1.2 percent) on the “stores” question and 3 respondents (.8 percent) on the “changes” question did not use the lines when provided. The number of respondents who left the “changes” question blank was significantly larger on the lined version (26.6 percent versus 19.4 percent;  $\chi^2 = 7.75, p = .005$ ). Respondents were more likely (though not significantly) to use prose on the version with the lines on the second question (48.0 percent versus 42.4 percent;  $\chi^2 = 2.4, p = .120$ ).

<sup>a</sup>  $N = 401$  for Lines, 407 for No lines; question provided a smaller answer space.

<sup>b</sup>  $N = 371$  for Lines, 394 for No lines; question provided a larger answer space.

**Table 6.** Independent Symbolic Language Manipulations: Percentage of Respondents Answering Subordinate Question, Arrow versus No Arrow

	Arrow (N=429)	No Arrow (N=434)
Mentions from Eligible	93.9	90.8
Mentions from Noneligible	2.1	.5
Missing (No response)	4.2	8.8
Total	100.0	100.0

NOTE.—Total mentions versus Missing:  $\chi^2 = 7.34$ ,  $p = .007$ . Eligible versus Noneligible mentions versus Missing:  $\chi^2 = 11.65$ ,  $p = .003$ . Eligible mentions versus Noneligible mentions:  $\chi^2 = 4.26$ ,  $p = .039$ .

This combined manipulation produced dramatic differences in the responses given by respondents. First, the use of the number box significantly increased the mean for each of the questions tested (table 7). The means varied from 2.4 to 2.8 for the first two questions and from 2.7 to 2.9 for the third question ( $p = .000$  for all three mean differences).

To search for an explanation for these observed differences, additional coding was conducted to see if respondents became confused and did not remember the direction of the scale (from positive to negative, consistent with other items in the same section). Responses to each of the three answer-box questions were coded based on whether there was any evidence of respondents changing their answers, and, as a control, evidence of changing answers on the polar point version were coded as well. To be included in the analysis, both the original and final answers must have been "legible," so any illegible answers were not coded (there were very few illegible answers and did not seem to vary by version). On the answer box version 10 percent (versus 1 percent on the polar point scales) of respondents scratched out answers to at least one of the questions and provided a different answer. A total of 74 respondents made 86 changes in their answers. Most of these errors occurred because respondents reversed the scale on the answer box version: 44 respondents changed their answers from 4 to 2; 10 respondents changed from 5 to 1; 4 changed from 2 to 4; and 1 changed from 1 to 5. These data suggest that removing the graphical layout of the scale from the polar point version affected respondents' understanding of the scale. Other respondents may have made this error without catching it, and this may be a cause of the larger mean scores in the answer box version.

To additionally test for respondent confusion, individual correlations were calculated between both versions of these 3 questions and 13 other questionnaire items about satisfaction with classroom experience, 12 of which were not varied across experimental treatments. Each of these items was expected

**Table 7.** Combined Graphical and Symbolic Language Manipulation: Percentage of Respondents Choosing Each Response on Polar Point Scale versus Answer Box

Responses	Question 1		Question 2		Question 3 <sup>a</sup>	
	Polar Point (N = 517)	Answer Box (N = 513)	Polar Point (N = 517)	Answer Box (N = 512)	Polar Point (N = 506)	Answer Box (N = 466)
(1) Very Satisfied	15.9	9.8	10.8	6.3	15.2	14.2
(2)	43.5	29.2	47.6	35.4	31.8	23.8
(3)	31.0	34.5	31.3	33.8	31.6	29.6
(4)	7.7	21.3	9.5	20.9	15.4	21.7
(5) Very Dissatisfied	1.9	5.3	0.8	3.7	5.9	10.7
Total	100.0	100.0	100.0	100.0	100.0	100.0
Mean	2.4	2.8	2.4	2.8	2.7	2.9
Difference of Means	$t = -7.7, p = .000$		$t = -6.9, p = .000$		$t = -3.5, p = .000$	
Chi-square	$\chi^2 = 63.4, p = .000$		$\chi^2 = 48.1, p = .000$		$\chi^2 = 18.0, p = .001$	

<sup>a</sup> Response categories for this item were (1) Outstanding to (5) Terrible.

to correlate positively with answers to these three test questions. If confusion existed among respondents and they did not change their answers, we would expect lower correlations between the 3 answer-box questions and the 13 other items, than for the polar point versions. All 13 correlations for each of the items using both formats were positive, but the mean correlation for the polar point format was .24, compared to only .14 for the answer box format. In only 4 of the 36 instances was the correlation between the 13 items and the test items higher for the answer box version. It was also observed that the mean difference for the correlations between versions for the first item (.16) was higher than the mean for the second (.10), with the third item reflecting the lowest mean difference (.03). This was consistent with the expectation that because of repetition respondents are less likely to make errors on the later items as they become accustomed to using the scale.

## Discussion and Conclusions

Most previous work in survey methodology has focused on question wording as the sole conduit of question meaning. In this article, we have proposed that the use of graphical and symbolic languages communicate additional meaning to the respondent and can independently and jointly influence respondent behavior.

These 14 experimental tests of 7 hypotheses combine to form an initial analysis of how graphical and symbolic languages influence respondent behavior in self-administered questionnaires. Significant differences were found for all of the hypotheses (though not on every experimental item), and most of the remaining tests were in the expected direction. Our general conclusion is that the visual design of questions on self-administered questionnaires can significantly impact respondent behavior.

Individual and combined manipulations of visual languages can influence how respondents navigate through a self-administered questionnaire as well as influence answers to specific survey questions. Instructions can vary in their impact, depending on whether they are placed in the respondent's navigational path or not. Directional arrows can help to guide respondents to subordinate questions. The amount of space provided in the questionnaire and how it is apportioned among response options can affect the way in which respondents choose answer categories and how much information they provide. Attempting to save space without careful consideration of the effects on how questions appear to respondents can lead to unintended consequences. Applying a response format suitable for one survey mode (answer boxes in telephone surveys) to another (the self-administered questionnaire) may lead to respondent confusion. The findings from this initial investigation highlight some of the possible effects of graphical and symbolic manipulation and provide a point of departure for further investigation.

Based on the experimental evidence presented here and in previous writings (Redline and Dillman 2002; Redline et al. 2003) as well as additional studies discussed earlier (Schwarz 1996; Schwarz et al. 1985, 1991; Schwarz, Grayson, and Knäuper 1998; Smith 1993, 1995), it is apparent that survey questions on self-administered questionnaires consist of much more than words. Respondents following the conduct of conversation actively construct meaning as they complete survey questions and gain meaning from the context of the questions, which includes the formal features of questionnaire design—symbols, numbers, and graphics. Future research needs not only to test the current hypotheses on other populations using different substantive items, but also needs to evaluate other manipulations, e.g., the role of numbers in getting people to answer questions in a prescribed order, the role of consistency in the use and display of symbols throughout questionnaires, the effects of figure/ground variations on whether information is processed by respondents, and how color and contrast can affect question comprehension. This additional research will aid in developing general principles about how questions should appear on questionnaire pages.

In addition, the increasing use of Internet surveys demands that these ideas be tested on Web questionnaires. Mixed-mode comparisons are also necessary to understand visual principles as they impact comparisons across self-administered and aural modes (telephone and face-to-face interviews). The recent trend toward greater use of mixed-mode surveys, in which researchers attempt to survey some members of a population via one mode and some by another, suggests that a priority for future research is to understand the fundamental causes of modal differences. Such work is essential for taking us beyond our current stage of understanding survey responses by verbal language alone to a better understanding of how verbal and nonverbal languages work together to convey meaning to respondents.

## References

- Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. 2d ed. New York: Wiley.
- Dillman, Don A., and Lisa Carley-Baxter. 1999. Unpublished data, Social and Economic Sciences Research Center, Washington State University, Pullman.
- Jenkins, Cleo R., and Don A. Dillman. 1997. "Towards a Theory of Self-Administered Questionnaire Design." In *Survey Measurement and Process Quality*, ed. Lyberg et al., pp. 165–96. New York: Wiley.
- Kahneman, D. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.
- Payne, S. L. 1951. *The Art of Asking Questions*. Princeton, NJ: Princeton University Press.
- Redline, Cleo, and Don A. Dillman. 2002. "The Influence of Alternative Visual Designs on Respondents' Performances with Branching Instructions in Self-Administered Questionnaires." In *Survey Nonresponse*, ed. R. Groves et al., pp. 179–93. New York: John Wiley and Sons.
- Redline, Cleo, Don A. Dillman, Aref Dajani, and Mary Ann Scaggs. 2003. "Navigational Performance in Census 2000: An Experiment of the Alteration of Visually Administered Languages." *Journal of Official Statistics* 19:403–19.

- Rothwell, Naomi D. 1985. "Laboratory and Field Response Research Studies for the 1980 Census of Population in the United States." *Journal of Official Statistics* 1:137-57.
- Sawyer, Scott and Don A. Dillman. 2002. *How Graphical, Numerical, and Verbal Languages Affect the Completion of the Gallup Q-12 on Self-Administered Questionnaires: Results from 22 Cognitive Interviews and a Field Experiment*. Social and Economic Sciences Research Center Technical Report no. 02-26, Pullman: Washington State University.
- Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.
- Schwarz, Norbert. 1996. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. Mahwah, NJ: Lawrence Erlbaum.
- Schwarz, Norbert, Carla E. Grayson, and Bärbel Knäuper. 1998. "Formal Features of Rating Scales and the Interpretation of Question Meaning." *International Journal of Public Opinion Research* 10(2):177-83.
- Schwarz, N., H. J. Hippler, B. Deutsch, and F. Strack. 1985. "Response Scales: Effects of Category Range on Reported Behavior and Subsequent Judgments." *Public Opinion Quarterly* 49:388-95.
- Schwarz, Norbert, Bärbel Knäuper, Hans J. Hippler, Elisabeth Noelle-Neumann, and Leslie Clarak. 1991. "Rating Scales: Numeric Values May Change the Meaning of Scale Labels." *Public Opinion Quarterly* 55:570-82.
- Sless, David. 1994. "Public Forums: Designing and Evaluating Forms in Larger Organizations." Paper presented at the International Symposium on Public Graphics, Lunteren, Netherlands.
- Smith, Tom W. 1993. *Little Things Matter: A Sampler of How Differences in Questionnaire Format Can Affect Survey Responses*. GSS Methodological Report no. 78. Chicago: National Opinion Research Center.
- . 1995. "Little Things Matter: A Sample of How Differences in Questionnaire Format Can Affect Survey Responses." Paper presented at the annual meeting of the American Association for Public Opinion Research, Fort Lauderdale, FL.
- Stember, Charles Herbert. 1956. "The Effect of Field Procedures on Public Opinion Data." Ph.D. dissertation, Columbia University.
- Sudman, S., and N. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine.
- Wright, P., and P. Barnard. 1975. "Just Fill in This Form—A Review for Designers." *Applied Ergonomics* 6:213-20.
- . 1978. "Asking Multiple Questions about Several Items: The Design of Matrix Structures on Application Forms." *Applied Ergonomics* 9:7-14.